

Gestió de dades de recerca: resultats de l'enquesta de 2018

(Doc. CO18/19) (4GDR\Enquesta\2018\1810 InformeNovaEdicioEnquesta_def.docx, 19.02.19)

Aquest document ha estat elaborat pel Grup de Treball de Suport a la Recerca del CSUC que està format per representants de les universitats següents: Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic-Universitat Central de Catalunya, Universitat Ramon Llull i Universitat Jaume I.

Aquest document està subjecte a la llicència de Reconeixement de Creative Commons (<http://creativecommons.org/licenses/by/4.0/>).

Versió digital: <http://hdl.handle.net/2072/351567>

1. Antecedents

El Grup de Treball de Suport a la Recerca del CSUC format per representats de la Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic-Universitat Central de Catalunya, Universitat Ramon Llull i Universitat Jaume I ha dut a terme una segona edició de l'enquesta que es va fer el 2016¹ sobre gestió de dades de recerca.

L'enquesta té la finalitat de conèixer les necessitats dels investigadors en matèria de gestió de dades de recerca així com tenir una visió general del grau de coneixement que es té de la matèria.

2. Mostra

Aquest informe ha estat elaborat en base a 709 respostes. El total per universitat és el següent: UB 86 (12%), UAB 182 (26%), UPC 117 (17%), UPF 14 (2%), UdG 24 (3%), UdL 77 (11%), URV 83 (12%), UOC 93 (13%) i URL 32 (5%). La UVic-UCC i la UJI no han enviat l'enquesta per motius interns.

L'enquesta ha estat elaborada en català, castellà i anglès i s'ha enviat de la forma que cada universitat ha considerat més convenient als col·lectius que cadascuna ha determinat entre el març i el juny de 2018. Hi ha universitats que només ho han enviat als investigadors principals dels projectes de recerca, en canvi d'altres l'han obert a doctorands i membres del personal d'administració i serveis (PAS). Aquest és un dels motius pels quals el nombre de respostes per universitat és tan desigual.

Donada la diversitat en els grups a qui s'ha enviat l'enquesta les dades de cada institució no són comparables. Es considera, però, que les dades serveixen per fer prospecció de les necessitats i seguir treballant en aquesta matèria.

3. Anàlisi dels resultats

S'han rebut 709 respostes. El 87,4% dels enquestats que han respost l'enquesta són Personal Docent i Investigador (PDI); en segon lloc, el col·lectiu de doctorands amb un 9% i, de manera residual, també han contestat membres del PAS (1,6%) i Altres (2%).

De les 709 respostes:

¹ Consorci de Serveis Universitaris de Catalunya. "Gestió de les dades de recerca: resultats de l'enquesta prospectiva a gener de 2016". 2016. <<http://hdl.handle.net/2072/268186>>. [Consulta: 31-10-18].

- 461 persones (65%) indiquen que tenen o han tingut algun tipus de relació amb la gestió de dades de recerca.
- 368 persones (35%) manifesten no haver tingut relació amb la gestió de dades de recerca.

Les dades d'aquesta enquesta es refereixen a aquestes 461 persones amb algun tipus de relació amb la gestió de dades de recerca, excepte la darrera pregunta que reflecteix la totalitat de les respostes (709).

De les 461 persones amb algun tipus de relació amb la gestió de dades de recerca:

- 328 (57%) han indicat que han compartit les seves dades voluntàriament.
- 133 (19%) han indicat que han participat o estan participant en algun projecte de recerca on s'ha demanat elaborar un pla de gestió de dades (PGD) i/o fer públiques les dades.

Hi ha 37 persones que afirmen haver fet PGD, però aquesta és una dada insegura perquè en alguns casos les respostes no són pertinents (per exemple, es confonen els PGD amb repositoris). Entre els que identifiquen l'eina usada per fer el Pla hi ha:

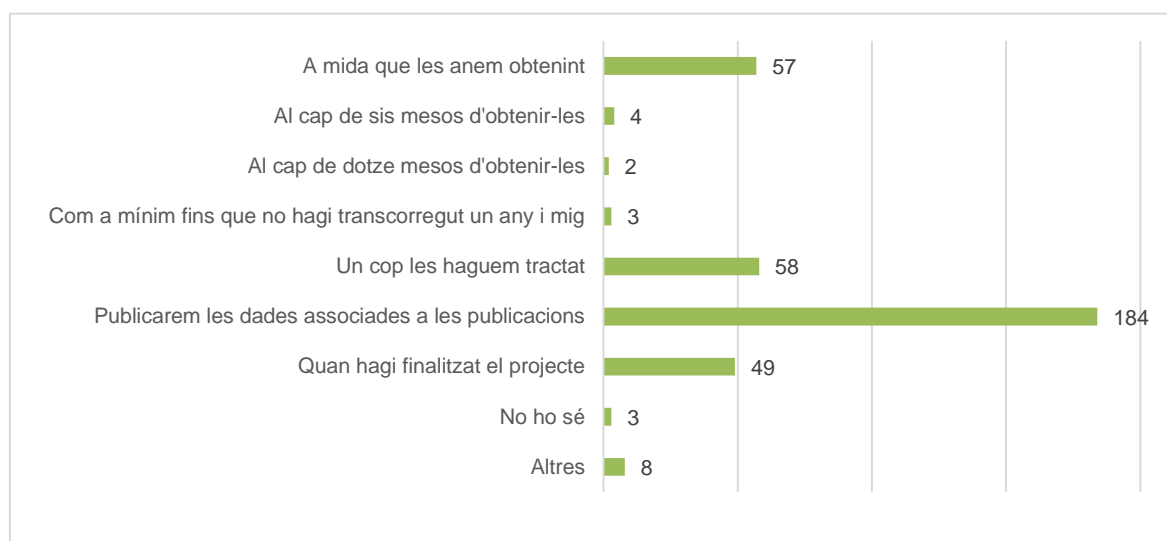
- 8, Research data management plan de les universitats catalanes
- 5, DMPOnline del DCC
- 3, DMPTool de la University of California Curation Center

Sobre les dades de recerca

A la pregunta sobre quin tipus de dades s'han generat, el 47% han indicat que dades observacionals, un 17% de simulació, i un 14% de referència i derivades o compilades.

Difusió de les dades

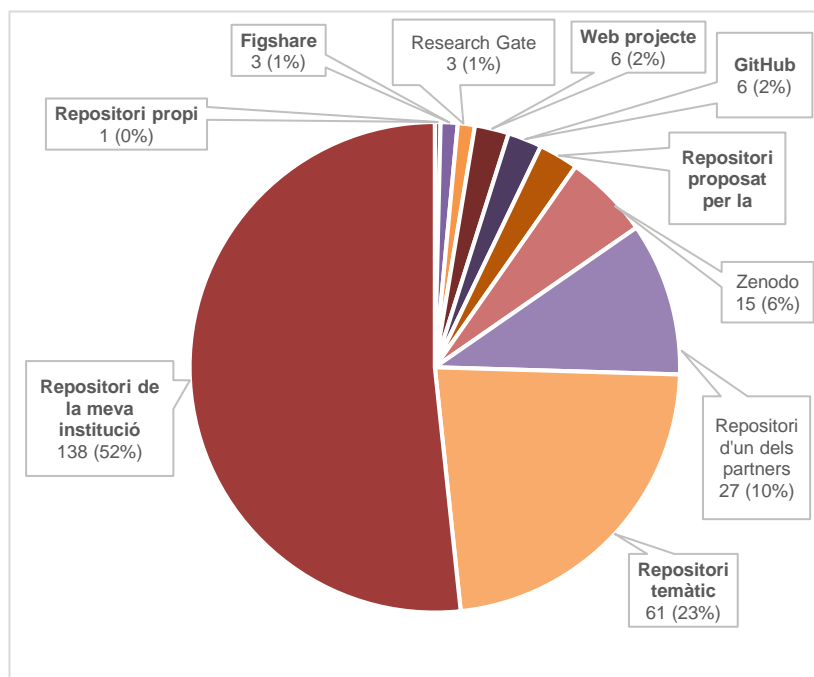
Un 60% dels enquestats han indicat que tenen previst dipositar i difondre les dades (totes o algunes) en obert. 184 (50%) volen fer-ho publicant les dades associades a les publicacions, 58 (15%) un cop s'hagin tractat les dades, 57 (15%) a mida que es vagin obtenint i 49 (13%) un cop finalitzi el projecte. Les altres opcions són minoritàries (vegeu gràfic 1).



Gràfic 1. Previsió sobre quan es publicaran les dades

D'altra banda, només un 11% diuen que no publicaran en obert i un 28% no n'està segur.

Malgrat que la majoria dels enquestats preveuen dipositar i difondre les dades en obert, el 47% desconeixen quina plataforma utilitzaran. De les opcions donades (vegeu gràfic 2), la que sobresurt és el repositori institucional (en menor presència, en repositoris temàtics o en el repositori d'un dels *partners*).

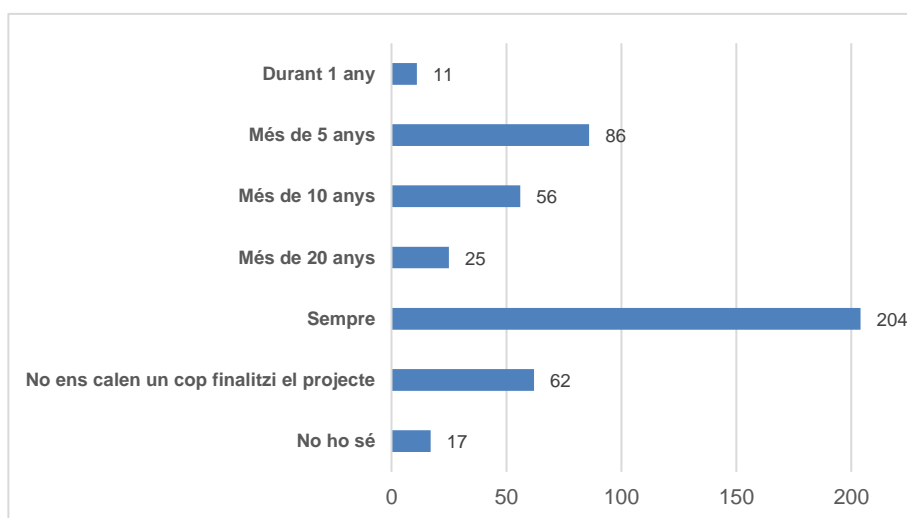


Gràfic 2. Plataforma on es publicaran les dades

Associat a la publicació de les dades, es pregunta sobre les llicències que es preveuen atorgar per a que puguin ser reutilitzades. El 65% dels enquestats no coneixen quina llicència atorgaran a les dades.

Preservació de les dades i ús d'estàndards

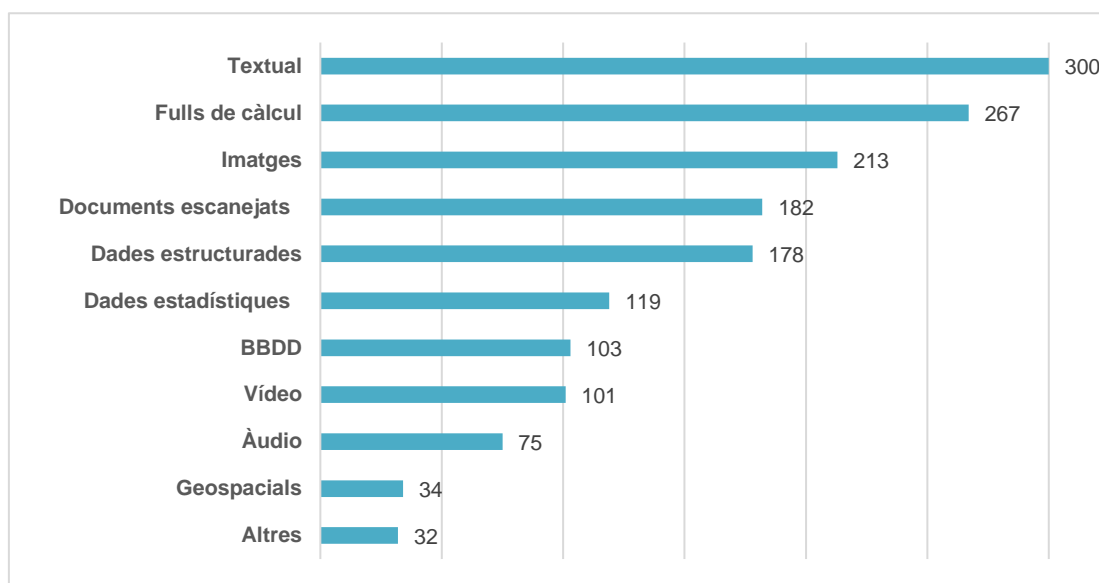
En general, els enquestats volen preservar les seves dades d'una manera o altra (vegeu gràfic 3). Gairebé la meitat (45%) indiquen que volen fer-ho per sempre. Només un 13% informen que no necessitaran fer-ho un cop finalitzi el projecte.



Gràfic 3. Temps estimat per preservar les dades

Aquestes dades a preservar tenen mides diverses però el 53% són inferiors a 10 GB.

La naturalesa de les dades amb les que treballen els investigadors és variable (vegeu gràfic 4). Pel que fa als formats dels fitxers generats en el marc del projecte de recerca que sobresurten són: el textual (18%), els fulls de càlcul (16%); en un segon bloc, les imatges (13%), els documents escanejats (11%) i les dades estructurades (11%) i, ja, en un últim bloc trobem d'altres formats com les dades estadístiques (7%), els vídeos (6%) o les bases de dades (6%), entre d'altres.



Gràfic 4. Format de dades amb les que es treballa

Dades de caràcter personal/confidencial

El 70% dels enquestats han respost que no treballen amb dades de caràcter personal, un 25% que sí i només un 5% no ho saben. Malgrat això, l'anonimització és la opció escollida en el cas que es vulguin publicar (el 63%).

Pel que fa a les dades de caràcter confidencial, el 78% indiquen que no treballen amb elles, el 18% sí i un 4% ho desconeixen. Dels que hi treballen, només el 19% publicaran les dades transcorregut un període temps i la resta, el 46% només les faran accessibles al grup de recerca i el 34% al projecte.

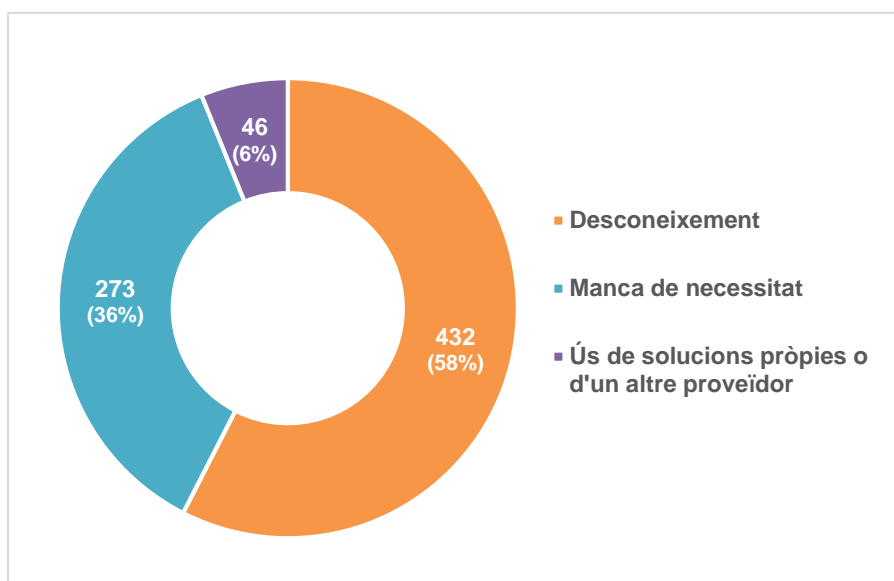
Responsabilitats en la gestió de les dades

La responsabilitat de la gestió de les dades en els projectes de recerca recau en el 60% dels casos en l'investigador principal del projecte. La resta d'opcions indicades, com són els grups o socis del projecte o el coordinador del projecte –entre d'altres–, tenen uns resultats molt minoritaris (menys de 10% en cada opció).

El servei de gestió de dades de recerca a la universitat

Des del setembre de 2016, les diferents universitats estan preparades per oferir suport en la gestió de dades de recerca i s'ha preguntat sobre aquest servei a tots els enquestats (709).

567 persones (80%) no coneixen el servei per diferents motius (vegeu gràfic 5): desconeixement (58%), manca de necessitat (36%) i ús de solucions pròpies o d'un altre proveïdor (6%).



Gràfic 5. Motius pels quals no es coneix el servei de suport a la gestió de dades de recerca de les universitats

Així mateix, només 64 persones (9%) l'han utilitzat en alguna ocasió per: poder dipositar les dades en obert al repositori (39%), demanar suport en drets d'autor i llicències (27%), citar les dades (23%) i elaborar un pla de gestió de dades (11%).

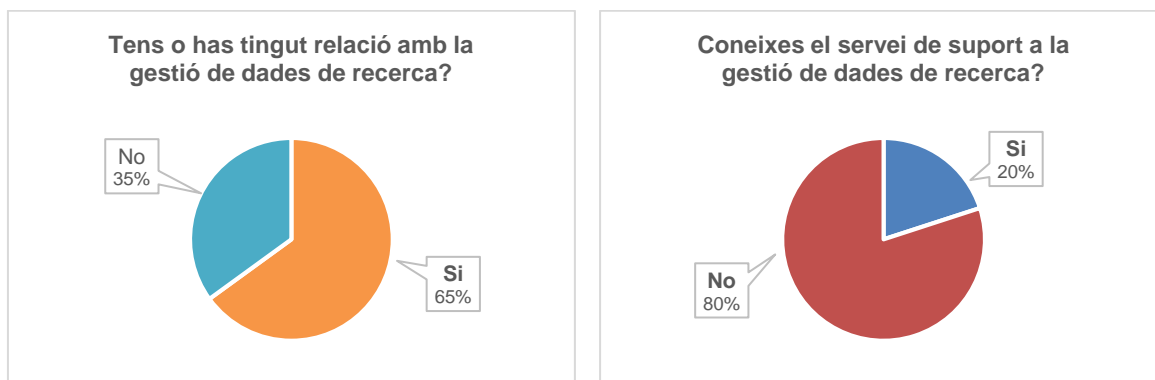
Comentaris

En aquesta edició, es va creure convenient incloure una pregunta oberta on els enquestats poguessin fer-hi aportacions addicionals. Entre els diferents comentaris rebuts, se'n destaca la necessitat de formació en aquesta matèria i conèixer quins serveis ofereix la universitat.

4. Conclusions

En relació amb el total dels enquestats:

- El 65% indiquen que tenen o han tingut alguna relació amb la gestió de dades de recerca.
- El 80% desconeixen el servei de suport a la gestió de dades de recerca que es dona a la seva universitat. Els motius donats són:
 - desconeixement (58%),
 - manca de necessitat (36%) o
 - ús de solucions pròpies o d'un altre proveïdor (6%).
- Només 64 persones (9%) han utilitzat el servei en alguna ocasió.



En relació amb els enquestats que indiquen que tenen o han tingut alguna relació amb la gestió de dades de recerca:

- El 60% tenen previst difondre les dades en obert.
- El 51% dels que volen publicar les dades, afirmen que ho faran al repositori institucional.
- El 65% desconeixen quin tipus de llicència atorgar a les dades.
- El 45% volen preservar les dades per sempre i el 13% no els calen preservar-les.
- Les dades són majoritàriament menors a 10GB i amb formats usuals.
- En el 60% dels casos, l'IP és el màxim responsable de la gestió de dades de recerca.

Annex

Gestió de dades de recerca (2a edició)

Benvolguts investigadors i benvolgudes investigadores,

En els darrers anys, les dades recollides, generades o utilitzades en la recerca estan rebent l'atenció de la comunitat científica i dels òrgans gestors de la recerca. La difusió d'aquestes dades és un dels elements de la ciència oberta.

Moltes entitats de finançament (Horizon 2020 de la Comissió Europea, el Plan Estatal 2017-2020, convocatòries d'Excel·lència María de Maeztu i Severo Ochoa, etc.) requereixen o promocien elaborar plans de gestió de dades i dipositar les dades en obert seguint els principis FAIR (Findable, Accessible, Interoperable i Reusable).

Les biblioteques i les oficines de recerca de les universitats catalanes -que treballen coordinadament en el si de l'Àrea de Ciència Oberta del CSUC- ofereixen serveis de suport a la gestió de dades de recerca.

L'objectiu d'aquesta enquesta és conèixer les teves necessitats per adaptar o reorientar aquests serveis.

Temps de resposta: entre cinc i deu minuts.

Gràcies per la teva col·laboració,

Dades personals

Institució *

- UB
- UAB
- UPC
- UPF
- UdG
- UdL
- URV
- UOC
- UVic-UCC
- URL
- UJI

1. Has participat o estàs participant en algun projecte de recerca on us hagin demanat elaborar un pla de gestió de dades i/o fer públiques les dades? *

- ☐ Sí (Vés a la secció 5 – Sobre el projecte de recerca)
- ☐ No (Vés a la pregunta 2)

2. Has compartit voluntàriament dades de recerca? *

- ☐ Sí (Vés a la secció 6 – Difusió de les dades)
- ☐ No (Vés a la secció 10 – El servei de gestió de dades de recerca a la universitat)

Sobre el projecte de recerca

Títol del projecte _____ *

3. El teu projecte va produir o produeix dades de recerca? *

- ☐ Sí
- ☐ No

4. Vas elaborar o estàs elaborant un Pla de Gestió de Dades (Data Management Plan, DMP)?

- ☐ Sí
- ☐ No

(En cas afirmatiu) Selecciona les eines utilitzades:

- ☐ Research Data Management Plan (dmp.csuc.cat)
- ☐ DMPonline (dmponline.dcc.ac.uk)
- ☐ DMPTool (dmptool.org)
- ☐ Altres:

5. Quin tipus de dades es van generar en el marc del projecte?

- ☐ Observacionals: dades capturades en temps real (p. ex., neuroimatges, dades de mostres, dades de sensors, dades d'enquesta)
- ☐ Experimentals: dades capturades en equips de laboratori (p. ex., seqüències de gens, cromatogrames, dades de camps magnètics)
- ☐ Simulació: dades generades a partir de models de prova (p. ex., climatològiques, matemàtiques o models econòmics)
- ☐ Derivades o compilades: dades reproduïbles però de difícil reproducció (p. ex., de text i mineria de dades, models 3D, bases de dades compilades)
- ☐ De referència: conglomerat o conjunt de dades (p. ex., bancs de dades de seqüències de gens, estructures químiques, o portals de dades espacials)
- ☐ Altres

Difusió de les dades

6. Tens previst el dipòsit i la difusió de les dades (totes o algunes) en obert?*

- ☐ Sí
- ☐ No
- ☐ No ho sé

(En cas afirmatiu) Quan preveus que es publiquin?

- ☐ A mida que les anem obtenint
- ☐ Un cop les haguem tractat
- ☐ Publicarem les dades associades a les publicacions
- ☐ Al cap de sis mesos d'obtenir-les
- ☐ Al cap de dotze mesos d'obtenir-les
- ☐ Com a mínim fins que no hagi transcorregut un any i mig
- ☐ Quan hagi finalitzat el projecte
- ☐ Altres:

7. Utilitzaràs alguna plataforma per dipositar i difondre dades?

- ☐ Sí
- ☐ No
- ☐ No ho sé

(En cas afirmatiu) Indica quina plataforma utilitzaràs

- ☐ Repositori institucional
- ☐ Repositori d'un dels partners
- ☐ Repositori temàtic
- ☐ Zenodo
- ☐ Figshare
- ☐ Dryad
- ☐ Altres:

8. Quina llicència utilitzaràs per publicar les dades?

- ☐ CC0 = Creative Commons Zero
- ☐ CCBY = Creative Commons Atribució
- ☐ Una altra llicència CC
- ☐ Una llicència específica per a dades
- ☐ No utilitzarem cap llicència, les publicarem amb tots els drets reservats
- ☐ No ho sé
- ☐ Altres:

Preservació de les dades i ús d'estàndards

9. Fes una breu descripció dels procediments que preveus per a la conservació de les dades més enllà de la vida del projecte, indicant:

Preservació de les dades?

- ☐ Volem preservar-les sempre
- ☐ Volem preservar-les més de 20 anys
- ☐ Volem preservar-les més de 10 anys
- ☐ Volem preservar-les més de 5 anys
- ☐ Volem preservar-les durant 1 any
- ☐ No ens cal preservar-les un cop finalitzat el projecte
- ☐ Altres:

Volum total de les dades que voldràs preservar?

- ☐ Més petit que 1 Mb
- ☐ Més petit que 1 Gb
- ☐ Més petit que 50 Gb
- ☐ Més petit que 100 Gb
- ☐ Més petit que 1 Tb
- ☐ Més petit que 50 Tb
- ☐ Més gran que 50 Tb
- ☐ Altres:

Apliques algun criteri a l'hora de seleccionar les dades a preservar?

- ☐ Sí
- ☐ No

(En cas afirmatiu) Quin? _____

Pots fer alguna estimació del cost de la preservació de les dades per any?

10. Quin és el format de les dades amb les que treballes:

- ☐ Textual (.doc, .docx, .log, .rtf, .txt...)
- ☐ Fulls de càlcul (.wks, .xls...)
- ☐ Documents escanejats (.pdf)
- ☐ Dades (.csv, .dat, .xml...)
- ☐ Imatges (.bmp, .gif, .jpg, .png, .psd...)
- ☐ Àudio (.mp3, .wav...)
- ☐ Vídeo (.avi, .mov, .mp4...)
- ☐ Dades estadístiques (.sav, .sdq, .spv...)
- ☐ Bases de dades (.db, .mbdb, .pdb, .sql...)
- ☐ Geoespacial (.kml, .gml, .wkt..)

11. Les dades amb què treballes precisen d'un programari específic per ser consultades?

- ☐ Sí
- ☐ No
- ☐ No ho sé

(En cas afirmatiu) De quin programari es tracta? _____

Dades de caràcter personal/confidencial

12. Treballes amb dades de caràcter personal?

- ☐ Sí
- ☐ No
- ☐ No ho sé

(En cas afirmatiu) Què faràs amb aquest tipus de dades?

- ☐ No les difondrem fora del nostre grup de recerca
- ☐ No les difondrem fora del nostre projecte
- ☐ Les publicarem anonimitzades
- ☐ Les publicarem agregades

13. Treballes amb dades confidencials (secret empresarial, patents, etc.)?

- ☐ Sí
- ☐ No
- ☐ No ho sé

(En cas afirmatiu) Què faràs amb aquest tipus de dades?

- ☐ Només seran accessibles al nostre grup de recerca
- ☐ Només seran accessibles al nostre projecte
- ☐ Les publicarem transcorregut un període de temps

Responsabilitats en la gestió de les dades

14. Qui és el màxim responsable de gestionar les dades?

- ☐ L'investigador principal del projecte
- ☐ El grup coordinador del projecte
- ☐ Un grup/soci específic del projecte
- ☐ Cada grup/soci del projecte
- ☐ Un soci extern contractat pel projecte
- ☐ No ho sé
- ☐ Ningú

El servei de gestió de dades de recerca a la universitat

15. Coneixes el servei de suport a la gestió de dades de recerca a la teva universitat?

- ☐ Sí
- ☐ No

16. L'has utilitzat en alguna ocasió?

- ☐ Si (Vés a la secció 11)
 - ☐ No (Vés a la secció 12)
-

Amb quina finalitat? (Secció 11)

- ☐ Elaborar un Pla de Gestió de Dades (DMP)
 - ☐ Dipositar en obert en un repositori
 - ☐ Suport en drets d'autor i llicències
 - ☐ Citar les dades
 - ☐ Altres:
-

Per quin motiu? (Secció 12)

- ☐ Desconeixement
 - ☐ Manca de necessitat
 - ☐ Ús de solucions pròpies o d'un altre proveïdor
 - ☐ Altres:
-

Comentaris

17. Aprofita i diga'ns la teva _____